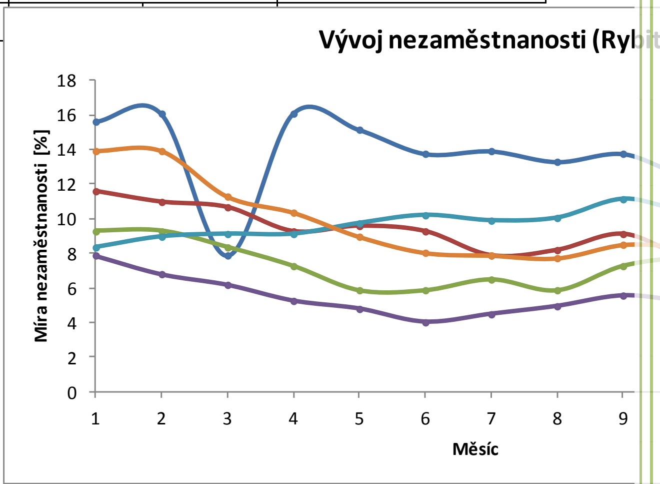
2011

Explorační analýza časových řad (Teorie)

Rok	y_t	3-členné klouzavé průměry	5-členné klouzavé průměry
1980	215		
1981	219	218,667	
1982	222	225,333	221,125
1983	235	219,667	218,000
1984	202	214,667	212,125
1985	207	198,667	203,875
1986	187	199,333	196,500
1987	204	188,333	188,625
1988	174	183,333	186,000
1989	172	182,333	196,250
1990	201	215	
1991	272		



Martina Litschmannová

VŠB-TU Ostrava

1 Explorační analýza časových řad

Co najdete v této kapitole?

- základní pojmy související s popisem časových řad,
- možnosti grafického zobrazení časových řad,
- způsoby průměrování časových řad,
- vyhlazování časových řad klouzavými průměry,
- základní míry dynamiky časových řad.

Časová řada je numerická proměnná, jejíž hodnoty podstatně závisí na čase, v němž byly získány (posloupnost chronologicky uspořádaných pozorování). Časové okamžiky, kdy byla data získána, jsou od sebe většinou stejně vzdáleny. Jde například o

- počty nezaměstnaných v jednotlivých měsících,
- počty automobilových nehod na Barandovském mostě v jednotlivých měsících,
- denní produkce mléka Veselé krávy.

Popis pomocí popisných statistik (krabicové grafy) – poskytuje dobrou představu o vlastnostech časové řady jako jednoho celku dat, ale neposkytuje informace o jejím časovém vývoji (roční průměrná mzda).

1.1 Základní pojmy

Časové řady lze klasifikovat podle různých hledisek, např.: podle charakteru dat, jejichž hodnoty tvoří časovou řadu

- **časové řady intervalové** - data závisí na délce intervalu, který je sledován (např. měsíční výroba cementu v ČR)
- **časové řady okamžikové** - data se vztahují k určitému okamžiku (počet nezaměstnaných v ČR v jednotlivých měsících)

podle periodicity, s jakou jsou data sledována

- **časové řady údajů ročních**
- **časové řady krátkodobé**

podle druhu sledovaných dat

- **časové řady absolutních ukazatelů** – např. počet obsloužených klientů za měsíc
- **časové řady odvozených charakteristik** – např. časová řada kumulativní (kumulativní časové řady, které vznikají postupným načítáním (kumulováním) jednotlivých hodnot (u

okamžikových časových řad nemají smysl, neboť výše jejich hodnot nezávisí na daném časovém intervalu, např. aktuální počet obsloužených klientů od začátku roku)

1.1.1 Očištění časové řady o důsledky kalendářních variací

Chceme-li porovnávat jednotlivé hodnoty u intervalových krátkodobých časových řad, musí se tyto hodnoty vztahovat ke stejně dlouhým časovým intervalům.

Očištění na měsíce

- standardní měsíc o délce 30 dnů – údaj za každý měsíc se vydělí počtem dnů v měsíci a vynásobí se 30, součet měsíčních údajů za rok potom odpovídá „roku“ o délce 360 dní

- standardní měsíc o délce 365/12 dnů – součet měsíčních údajů za rok odpovídá délce roku 365 dní

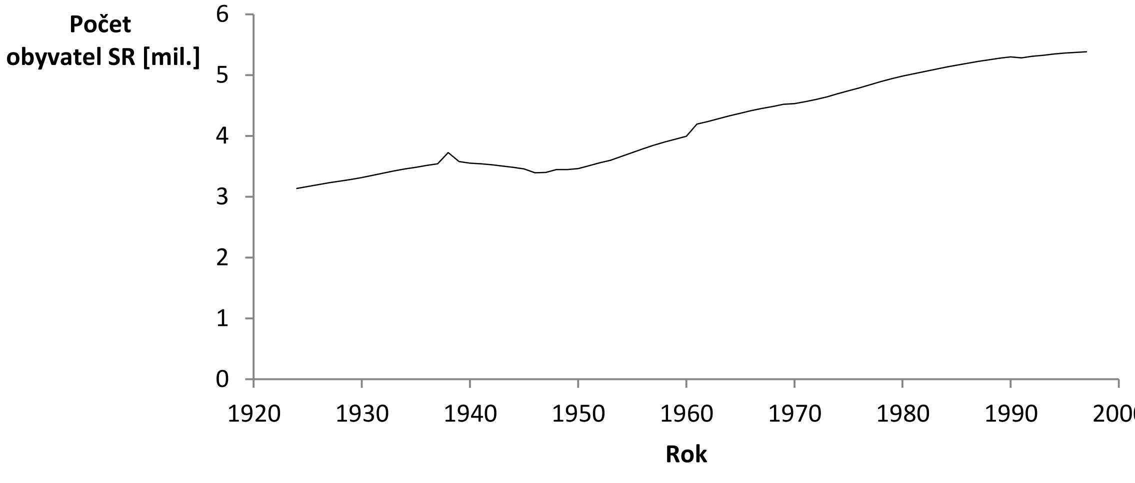
očištění na pracovní dny – provádí se obdobně jako očištění na měsíce

1.2 Grafická analýza časových řad

Jedním ze základních prostředků prezentace časových řad je jejich graf. Nejčastěji se graficky znázorňují původní hodnoty časové řady, nebo řady kumulativní. Často se ale časové řady zobrazují tak, aby více vynikly jejich charakteristické vlastnosti a rysy. K tomu slouží speciální typy grafů.

1.2.1 Spojnicový graf jedné časové řady

Základní informace pro analýzu časových řad získáme ze spojnicových grafů. Jejich princip spočívá v zakreslení jednotlivých hodnot časové řady do souřadnicového systému. Na osu horizontální se vynáší časová proměnná a na osu vertikální hodnoty časové řady nebo její funkce.



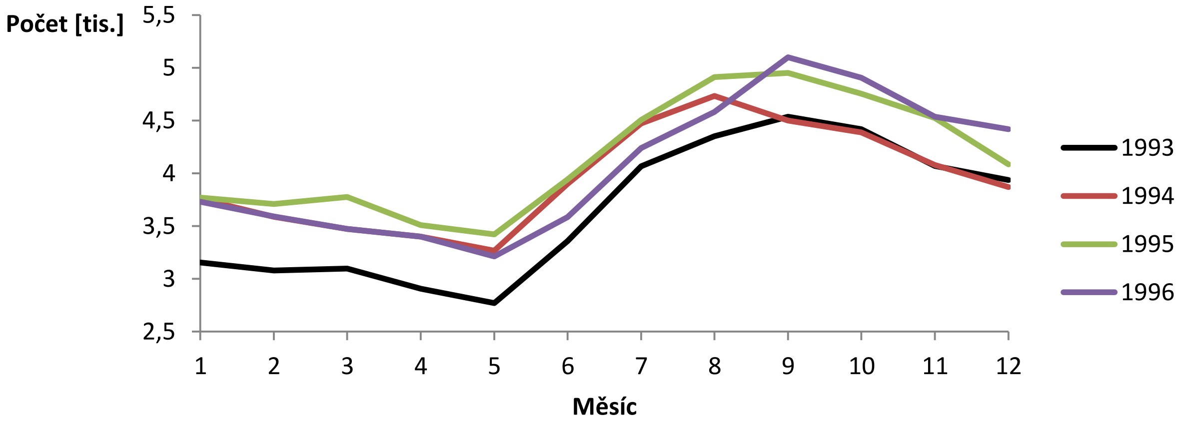
Obr. 1.1: Vývoj počtu obyvatel SR v letech 1924-1997

1.2.2 Spojnicový graf dvou a více časových řad

Do spojnicového grafu můžeme zakreslit i více časových řad. V případě, že zobrazujeme např. dvě časové řady lišící se měřítkem, je možné použít kromě levé i pravou vertikální osu.

1.2.3 Graf ročních hodnot sezónních časových řad

Speciálním případem spojnicového grafu dvou a více časových řad je graf ročních hodnot sezónních časových řad. Tento graf zobrazuje hodnoty časové řady uspořádané podle roků a tak charakterizuje, jak se v jednotlivých letech liší úroveň hodnot v daných sezónách za celou časovou řadu.



Obr. 1.2: Vývoj počtu nezaměstnaných absolventů gymnázií v SR

1.3 Popisné charakteristiky časových řad

1.3.1 Průměrování časových řad

Při práci s časovými řadami je někdy důležité zjistit jejich průměrné hodnoty.

- **intervalové řady** – výpočet se provádí pomocí aritmetického průměru.
- **okamžikové řady** – výpočet se provádí v případě stejných vzdálenosti mezi jednotlivými okamžiky pomocí prostého chronologického průměru, v případě nestejných vzdálenosti pomocí váženého chronologického průměru

Nechť časovým okamžikům $t_1, t_2, \ldots, t_n$ odpovídají hodnoty časové řady $y_1, y_2 \ldots, y_n$:

Prostý chronologický průměr: $$\frac{\frac{y_1}{2} + y_2 + \ldots + y_{n-1} + \frac{y_n}{2}}{n-1}$$

Vážený chronologický průměr:

$$\frac{\frac{y_1 + y_2}{2}(t_2 - t_1) + \frac{y_2 + y_3}{2}(t_3 - t_2) + \cdots + \frac{y_{n-1} + y_n}{2}(t_n - t_{n-1})}{t_n - t_{n-1}}$$

1.3.2 Míry dynamiky

Kromě průměrů nás mnohdy zajímají základní míry dynamiky časových řad, které umožňují charakterizovat základní rysy jejich "chování". Charakteristiky, které si dále uvedeme vyžadují stejnou délku časových intervalů v intervalových časových řadách nebo stejné vzdálenosti mezi okamžiky zjišťování v okamžikových časových řadách.

- **Absolutní přírůstky**

  Nejjednodušší mírou dynamiky je absolutní přírůstek, který nám říká „o kolik" se změnila časová řada mezi jednotlivými okamžiky.

$$\Delta^{(1)}y_t = y_t - y_{t-1}, \quad t = 2,3,\dots,n$$

- **Průměrný absolutní přírůstek**

  nám říká „o kolik" se průměrně změnila časová řada za období mezi dvěma měřeními během sledovaného období.

$$\bar{\Delta} = \frac{1}{n-1}\sum_{t=2}^{n}\Delta^{(1)}y_t = \frac{y_n - y_1}{n-1}$$

- **Koeficienty růstu**

  Koeficienty růstu udávají „kolikrát" se změnila časová řada mezi jednotlivými okamžiky.

$$k_t = \frac{y_t}{y_{t-1}} \qquad t = 2,3,\dots,n$$

- **Průměrný koeficient růstu**

  nám říká „kolikrát" se průměrně změnila časová řada za období mezi dvěma měřeními během sledovaného období. Vzhledem k tomu, že průměrujeme poměrové proměnné, musíme pro jeho výpočet použít geometrický průměr.

$$\bar{k} = \sqrt[n-1]{k_2 k_3 \dots k_n} = \sqrt[n-1]{\frac{y_n}{y_1}}$$

- **Meziroční koeficienty růstu**

  jsou podíly hodnot časové řady ve stejných obdobích (sezónách) v po sobě jdoucích letech. V případě čtvrtletní časové řady má meziroční koeficient růstu tvar

$$k_t = \frac{y_t}{y_{t-4}}, \qquad t = 5,6,\dots,n.$$

- **Relativní přírůstky [%]**

  Chceme-li vědět „o kolik procent" se změnila časová řada mezi jednotlivými okamžiky, použijeme relativní přírůstky.

$$\delta_t = \frac{\Delta^{(1)}y_t}{y_{t-1}}\cdot 100 = \frac{y_t - y_{t-1}}{y_{t-1}}\cdot 100 = (k_t - 1)\cdot 100, \qquad t = 2,3,\dots,n$$

- **Průměrný relativní přírůstek [%]**

udávající „o kolik %“ se průměrně změnila časová řada za období mezi dvěma měřeními během sledovaného období pak jednoduše vypočteme dle vztahu

$$\bar{\delta}_t = (\bar{k} - 1) \cdot 100\ [\%].$$

1.4 Dekompozice časových řad

Časovou řadu lze rozložit na součet (nebo součin) několika složek, z nichž každá bude podstatně jednodušší a bude mít jasnou interpretaci. Těmito složkami jsou:

- Trend D_t
- Sezónní složka S_t
- Cyklická složka C_t
- Náhodná složka E_t

Trend

- Odráží dlouhodobý vývoj (obvykle růst nebo pokles, ale obecně nemusí být tato složka monotónní) daného procesu.
- Co je dlouhodobé? (růst teploty během dne z pohledu lyžaře, zemědělce, meteorologa).

Sezónní složka

- Odráží periodické změny, které se mohou v dané řadě projevovat, a jejich perioda je svázána s kalendářem (mají periodu jednu hodinu, jeden den, týden, měsíc, rok, století…).
- Velký význam v ekonomii, meteorologii…

Cyklická složka

- Odráží periodické změny, které se mohou v dané řadě projevovat, a jejichž perioda neodpovídá délce nějaké kalendářní jednotky.
- V technických vědách se sezónní složka obvykle neuvažuje a všechny periodické jevy se zahrnují do cyklické složky.

Náhodná (reziduální) složka

- Zbývá v časové řadě po odstranění trendu, sezónních a cyklických složek.
- Je tvořena náhodnými fluktuacemi, které nemají žádný systematický charakter.

Nadále budeme pracovat vždy s náhodným procesem a s některou jeho realizací. Pro odlišení budeme časové řady označovat velkými písmeny (např. X_t) a jejich konkrétní realizace malými písmeny (např. x_t).
Znalost každé jednotlivé složky nám umožní například lepší odhad vývoje daného procesu do budoucna (predikci).

Budeme-li se snažit rozložit časovou řadu $\{X_t\}$ na součet složek, budeme předpokládat, že ho lze zapsat ve tvaru:

$$X_t = D_t + S_t + C_t + E_t$$

Tomuto způsobu rozkladu časové řady říkáme aditivní rozklad a používáme jej v případě, že se variabilita hodnot časové řady se v průběhu času příliš nemění.

Mění-li se variabilita hodnot časové řady v čase výrazně, používáme tzv. multiplikativní model.

$$X_t = D_t \cdot S_t \cdot C_t \cdot E_t$$

1.4.1 Metody hledání trendu

Pro hledání trendu se používají buď regresní metody (předpokládáme, že $X_t = D_t + E_t$), nebo adaptivní přístupy, které dokážou reagovat na změny v charakteru trendu. Mezi nejznámější adaptivní přístupy patří metoda klouzavých průměrů.

- **Klouzavé průměry**

Chceme-li z časové řady odstranit šum vznikající působením náhodných vlivů, lze použít metodu klouzavých průměrů spočívající v tom, že se řada původních pozorování nahradí řadou vypočtených klouzavých průměrů. Tato metoda je adaptivní, tzn. dokáže pracovat s časovou řadou, která v čase mění svůj charakter a nelze ji proto popsat jedinou křivkou.

Klouzavý průměr je určitou lineární kombinací *2p+1* členů původní řady. Čím větší je délka klouzavého průměru, tím větší je „vyhlazení" časové řady. V případě, že zvolená délka klouzavého průměru je „lichá", získáme jejich hodnoty jako obyčejné aritmetické průměry dané délky (prosté klouzavé průměry).

Prosté klouzavé průměry

Úseky časové řady o délce *2p+1* vyrovnáme tak, že je nahradíme prostým aritmetickým průměrem.

$$\overline{y_t} = \frac{1}{2p+1}\sum_{i=-p}^{p} y_{t+i} = \frac{y_{t-p} + y_{t-p+1} + \ldots + y_{t+p-1} + y_{t+p}}{2p+1} \qquad t = p+1, p+2, \ldots, n-p$$

p hodnot na začátku a p hodnot na konci časové řady zůstává nevyrovnáno.

Sudá délka klouzavých průměrů se volí jen velmi zřídka. V případě, že délku klouzavých průměrů zvolíme rovnu *2p*, používáme tzv. centrovaných klouzavých průměrů.

Centrované klouzavé průměry

$$y_t = \frac{1}{4p}\left(y_{t-p} + 2y_{t-p+1} + \cdots + 2y_{t+p-1} + y_{t+p}\right) \quad t = p+1, \ldots, n-p$$

Myšlenka tohoto vyrovnání je prostá. K tomu, aby vyrovnaná hodnota odpovídala danému období, potřebujeme lichý počet členů v klouzavém průměru. Ten získáme nejlépe tak, že místo prvního členu v klouzavém průměru vezmeme průměr první a poslední hodnoty dané periody.Tedy klouzavé průměry mají tvar

$$y_t = \frac{\frac{y_{t-p} + y_{t+p}}{2} + y_{t-p+1} + \cdots + y_{t+p+1}}{2p} = \frac{1}{4p}\left(y_{t-p} + 2y_{t-p+1} + \cdots + 2y_{t+p-1} + y_{t+p}\right).$$

U neperiodických časových řad se nejčastěji používají průměry délky 3, 5, 7. Pokud chceme z naší časové řady odstranit sezónní vliv (např. kolísání hodnot během týdne, měsíce…), využíváme klouzavých průměrů s délkou rovnou délce sezónního období.

Příklad: Časová řada (viz. tab.) udává roční objemy vývozu piv (v mil. l.) z ČSFR v letech 1980 až 1991. Vyrovnejte časovou řadu 3-člennými a 4-člennými klouzavými průměry.

Řešení:

Rok	y_t	**3-členné klouzavé průměry**	**5-členné klouzavé průměry**
1980	215		
1981	219	218,667	
1982	222	225,333	221,125
1983	235	219,667	218,000
1984	202	214,667	212,125
1985	207	198,667	203,875
1986	187	199,333	196,500
1987	204	188,333	188,625
1988	174	183,333	186,000
1989	172	182,333	196,250
1990	201	215	
1991	272		

3-člennéklouz. průměry: $2p + 1 = 3 \;\Rightarrow p = 1 \;\Rightarrow$ 1 hodnota na začátku a 1 hodnota na konci bude vynechána.

3-členný klouzavý průměr v roce 1981 = $\dfrac{215 + 219 + 222}{3} = 218{,}667$

3-členný klouzavý průměr v roce 1982 = $\dfrac{219 + 222 + 235}{3} = 225{,}333$

4-členné klouz. průměry: $2p = 4 \;\Rightarrow p = 2 \;\Rightarrow$ 2 hodnoty na začátku a 2 hodnoty na konci budou vynechány.

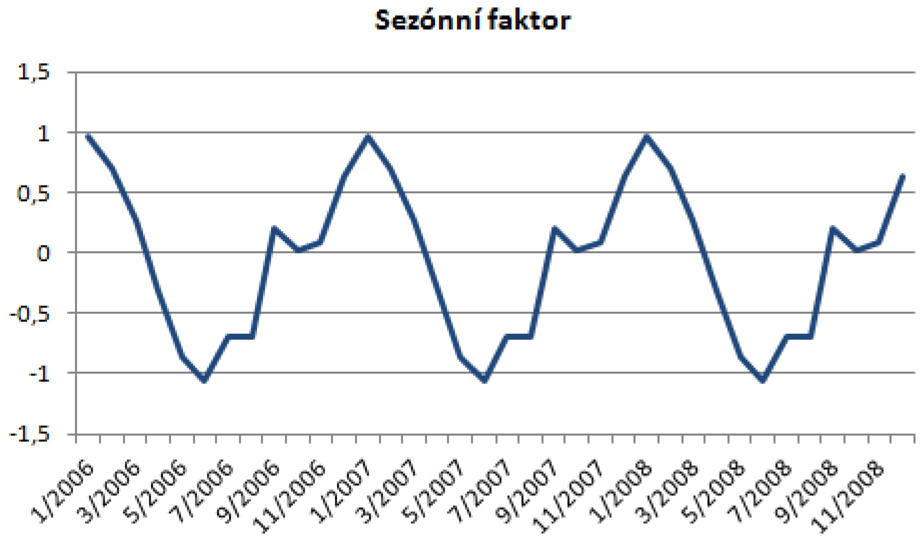
4-členný klouzavý průměr v roce 1982 = $\dfrac{\frac{215}{2}+219+222+235+\frac{202}{2}}{4} = 221{,}125$

atd.

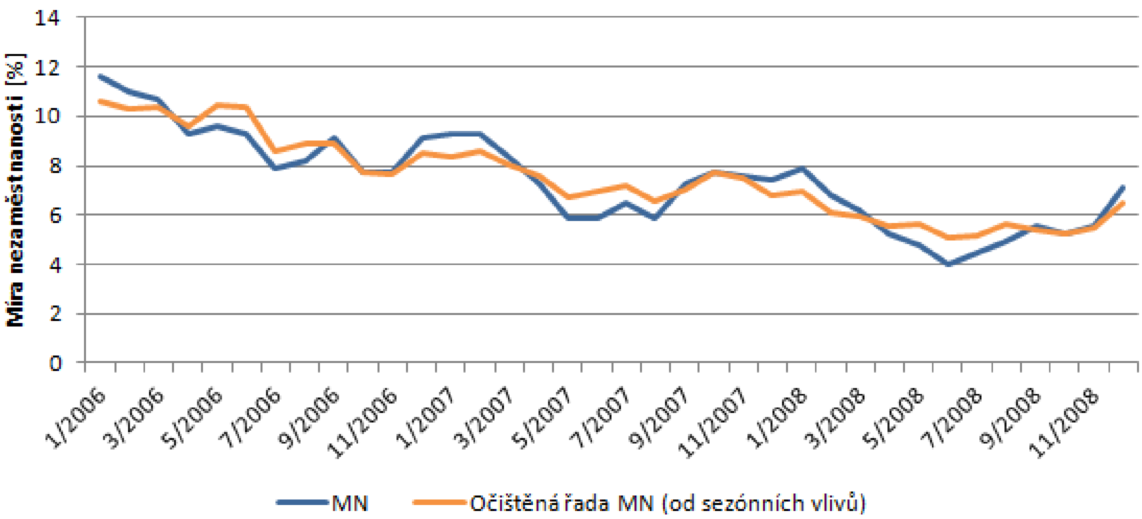
1.4.2 Očištění časové řady od sezónních vlivů

Pro očištění časové řady musíme nejdříve stanovit sezónní faktor v aditivní formě [2]. **Sezónní faktor** stanovíme pomocí **odchylky** časové řady a centrovaných klouzavých průměru o délce rovné periodě časové řady, nejčastěji o délce 12). (viz [2], příklad 6.4)

Sezónní faktor pro určitý měsíc pak určíme jako průměrnou měsíční odchylku, tj. lednový sezónní faktor se určí jako průměr všech lednových odchylek. (Všimněte si, že sezónní faktor je pro všechny roky stejný – viz obrázek.)



Časovou řadu očištěnou od sezónní složky získáme tak, že sezónní faktor odečteme od původní časové řady. Takto očištěná časová řada se pak používá pro další statistické vyhodnocení (regresní analýza, modelování časových řad, atd. [2]).



Literatura

[1] LITSCHMANNOVÁ, M. (2011), *Úvod do statistiky*, skripta - pilotní verze , dostupné z: http://mi21.vsb.cz/modul/uvod-do-statistiky
[2] HANČLOVÁ, J., TVRDÝ, L., Úvod do analýzy časových řad, dostupné z: http://gis.vsb.cz/pan/cz/skoleni.html